R-D Optimized Quantization of H.264 SP-Frames for Bistream Switching under Storage Constraints

Chen-Po Chang and Chia-Wen Lin
Department of Computer Science & Information Engineering
National Chung Cheng University
Chiayi 621, Taiwan
cwlin@cs.ccu.edu.tw

***Abstract*—In this paper, we propose a two-pass R-D optimized quantization scheme for improving the coding performance of H.264 primary SP-frames under a given storage constraint on secondary SP-frames. The first-pass encoding of the proposed scheme collects the statistics required for fitting the model parameters which are used to characterize the bit-rates of secondary SP-frames and the distortion of primary SP-frames. As a result, in the second-pass encoding, these estimated model parameters are then used to obtain the optimal set of quantization parameters for all SP-frames of each GOP using the proposed R-D optimized quantization schemes. Experimental results show that our proposed schemes can improve coding performance of primary SP-frames under the given rate constraint.**

I. INTRODUCTION

With the proliferation of online multimedia contents, the popularity of multimedia streaming technologies, and the establishment of video coding standards, people are able to ubiquitously access and retrieve various multimedia contents via the Internet, promoting networked multimedia services at an extremely fast pace. In streaming video, users may access videos from heterogeneous networks such as Local Access Network (LAN), Digital Subscriber Line (DSL), Cable, wireless networks, and dial-up. The different access networks have different channel characteristics such as bandwidths, bit error-rates, and packet loss-rates. At the users' end, network appliances including handheld computers, Personal Digital Assistants (PDA), set-top boxes, and smart cellular phones are slated to replace personal computers as the dominant terminals for accessing the Internet. These network terminals vary significantly in resources such as computing power and display capability. To flexibly deliver multimedia data to users with different available resources, access networks, and interests, the multimedia contents may need to be adapted dynamically according to the usage environment.

There are some traditional methods for video adaptation in a heterogeneous environment [1]. One is to encode the bitstream at a highest bit-rate/resolution of the Internet then transcodes the bitstream into different bit-rates/formats. First, the transcoder decodes the encoded bitstream, and then re-encodes it to meet the bit-rate and/or resolution that is suitable for each client. In this way, the streaming video provider can use a transcoder to transcode the bitstream into different bit-rates, resolutions, and formats for different users. But the transcoder may require large computing power and time cost to transcode. Another is to encode a scalable bitstream. A general scalable bitstream contains one base-layer and one or more enhancement-layers. Because temporal predictions are applied in the base-layer only with minimum perceptual quality, the scalable bitstream alone may not provide a large enough bit rate range to address large bandwidth variation without sacrificing the coding efficiency.

Dynamic bitstream switching [2] is another efficient means which has been widely deployed in commercial streaming services to deal with bandwidth variation in a standard compliant way. With bitstream switching, the server provides multiple bitstreams with different bitrates/resolutions for each client to switch over the bitstreams to choose the bitstream which matches the client's channel bandwidth the most for rate adaptation. For instance, clients with high channel bandwidths can subscribe to higher-rate bitstreams for better video quality, whereas low-bandwidth clients need to subscribe to lower-rate bitstreams with worse perceptual visual qualities. There are some issues with bitstream switching schemes to concern about. For example, when the available channel bandwidth of a client drops, the client has to switch from one higher-rate bitstream to another lower-rate one (a "switching-down" process), and vice versa (a "switching-up" process). Because general video coding schemes use the temporal predictive coding, switching at any predictive frame would cause different reference frames at the encoder and the decoder. This mismatch leads to drift which will propagate to subsequent predictive frames until reaching the next intra frame [2].

In order to mitigate the quality drift caused by bitstream switching, a pioneering work in [2] proposes to use a new-type intermediate switching frame (S-frame) to compensate for the switching drift at predictive frames. The S-frames can effectively reduce the switching drift but cannot eliminate the drift completely if they are not coded losslessly. Other attempts of controlling the quality drift are to dynamically choosing proper switching points (e.g., frames in a stationary neighborhood or in a neighborhood with the highest coding quality) based on source characteristics to achieve graceful transition while switching [3,4]. Recently, the H.264 standard has proposed a new picture type, the SP-frames [5,6], which supports drift-free switching at predictive frames. Like normal predictive frames (P-frames), SP-frames use motion compensated predictive coding to remove the temporal redundancy, while allowing identical reconstruction of the frames at switching points even when they are predicted with different reference frames [5,6]. The SP-frames can provide seamless switching points just like intra frames, but their frame sizes are much smaller than intra frame due to the predictive coding. However, the bit-rate required for encoding an SP-frame is still significantly higher than that for a normal P-frame [5].

0-7803-8834-8/05/$20.00 ©2005 IEEE.

II. BITSREAM SWITCHING WITH H.264 SP-FRAMES

The H.264 SP-frames, similar to normal P-frames, also adopt motion-compensated predictive coding for reducing the temporal redundancy between consecutive frames, but they allow switching from one bitstream to another one of different bit-rate/resolution without introducing any drift. Suppose one sequence is encoded into two bitstreams with different bit-rates. Fig. 1 illustrates an example of using an SP-frame to switch form one bitstream to another. As shown in Fig. 1, the SP-frames can be classified into primary SP-frames (e.g., S_1 and S_2) and secondary SP-frames (e.g., S_{12}), respectively. The secondary SP-frames are the special frames which can be used for switching up or down without drift just like switching at intra frames. They are transmitted while switching between two bitstreams. For example, in Fig. 1, if the server wants to switch from bitstream 1 to bitstream 2, it sends S_{12} instead of S_1 or S_2 to the decoder at the switching point.

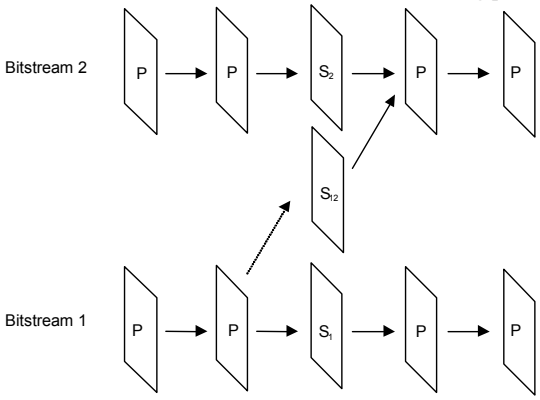


Fig. 1. Illustration of bitstream switching using SP-frames.

Fig. 2 depicts the encoder block diagram for generating H.264 primary SP-frames [5]. Compared to the P-frame encoding process, the primary SP-frame encoding process involves an extra re-quantization (also followed by an inverse quantization) process with the quantization step-size of the corresponding secondary SP-frame (Q_s). Using this additional quantization process, for the example shown in Fig. 1, the reconstructed S_{12} frame can be exactly identical to the reconstructed S_2 frame, thereby achieving seamless bitstream switching [5,6] without introducing mismatch error between S_{12} and S_2.

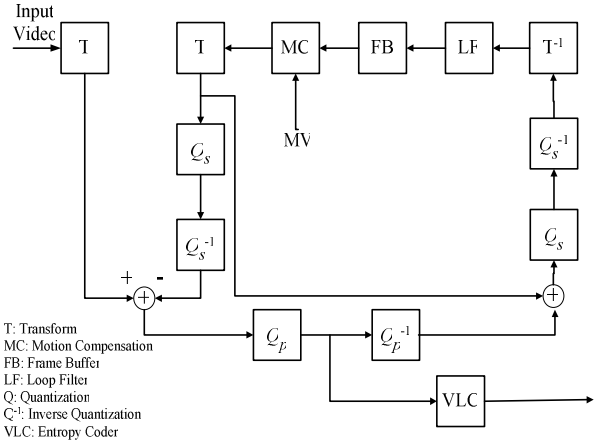


Fig. 2. Block diagram of H.264 primary SP-frame encoder.

Although the extra re-quantization of SP-frames with Q_s can achieve drift-free bitstream switching, it will lead to coding performance degradation of primary SP-frames as shown in Fig. 3, where the distortion is measured by SSD (Sum of Square Error). The coarser the quantization step-size (Q_s), the more significant the coding distortion. When $Q_s = Q_p$ (Q_p = 22), the distortion of primary SP-frames is about 1.5 times than that of P-frames (Q_p = 1). The PSNR performance degradation can be up to 3~4 dB when $Q_s = Q_p$ and Q_p is set relatively large for low bit-rate applications.

From Fig. 3, we can observe that the Q_s values not only influence the coding performance of primary SP-frames, but also affect the bit-rates of secondary SP-frames. Because the bit-counts of secondary SP-frames are usually much higher (3~4 times for a small Q_s) than those of normal P-frames, a significant amount of additional space is required for storing the extra SP-frames. Therefore, it becomes a trade-off between the coding performance of primary SP-frames and the storage cost for secondary SP-frames, which is similar to the problem addressed in [7]. For example, if we want to reduce the requantization distortion of a primary SP-frame, its Q_s value has to be very small, leading to a significantly high bit-count for the secondary SP-frame.

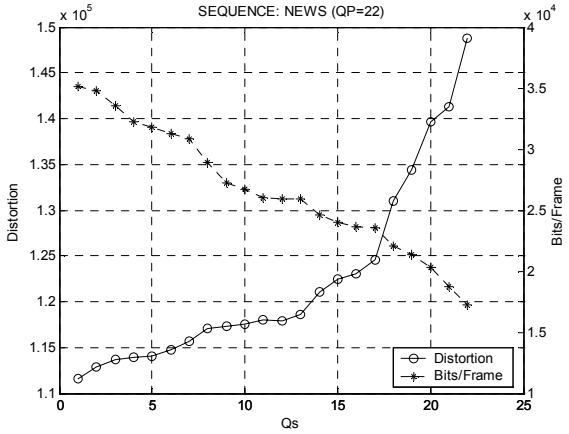


Fig. 3. Average bit-count of secondary SP-frames and SSD distortion of primary SP-frames for different Q_s values. (News; Q_p = 22)

III. R-D OPTIMIZED QUANTIZATION FOR SP-FRAMES

As mentioned above, the selection of the extra re-quantization step-sizes (Q_s) directly influences the bit-rate of secondary SP-frames, as well as the coding performances of the corresponding primary SP-frames and the subsequent P-frames within the same GOP. According to the R-D characteristics between the distortions of primary SP-frames and the bit-rates of secondary SP-frames, we propose a scheme of finding the optimal combination of Q_s values that minimizes the overall distortion of primary SP-frames subject to a reasonable storage constraint posed on secondary SP-frames as formulated in (1).

$$\min_{Q_{s,i}} \sum_{i=2}^{N_{GOP}} D_i(Q_{s,i}) \quad \text{subject to} \sum_{i=2}^{N_{GOP}} R_i(Q_{s,i}) \le R_C = \alpha R_H \qquad (1)$$

where N_{GOP} represents the GOP size, $Q_{s,i}$ is the re-quantization step-size of the ith frame in the GOP, R_C is the target storage constraint for secondary SP-frames, R_H is the bit-rate of the higher bit-rate bitstream, and α is a constraint factor.

Eq. (1) contains two rate and distortion functions: $D_i(Q_{s,i})$ and $R_i(Q_{s,i})$, where $D_i(Q_{s,i})$ represents the reconstruction distortion function for the ith primary SP-frame coded with $Q_{s,i}$, and $R_i(Q_{s,i})$ represents the bit-count of the ith secondary SP-frame as a function of $Q_{s,i}$. Therefore (1) indicates that we would minimize the overall distortion of primary SP-frames under the rate constraint R_C. Note that, the selection of Q_s only has a very minor impact (≤ 6%) on the bit-rates of primary SP-frames, which is thus negligible in (1).

In order to solve the constrained optimization problem in (1)

analytically, we have to find proper mathematical models for $D(Q_s)$ and $R(Q_s)$ with good accuracy. In general, the distortion of encoding a secondary SP-frame can be further divided into two components: the re-quantization distortion $D_i^{\mathrm{rq}}(Q_{s,i})$ and the propagated distortion (due to its reference picture) D_i^{ref}, where $D_i^{\mathrm{rq}}(Q_{s,i})$ represents the distortion caused by the extra re-quantization process itself, which is only affected by the re-quantization step-size $Q_{s,i}$; whereas D_i^{ref} accounts for the error propagation to the subsequent frames within the GOP caused by the re-quantization distortion due to the temporal prediction used in the primary SP-frames. We use the following function to represent the combined distortion:

$$D_i(Q_{s,i}) = D_i^{\mathrm{rq}}(Q_{s,i}) + D_i^{\mathrm{ref}} \cong (1+\mu_{\mathrm{ref}})^{N_{\mathrm{GOP}}-i} D_i^{\mathrm{rq}}(Q_{s,i}) \quad (2)$$

where μ_{ref} is an error propagation factor, which is related to the average Q_s in a GOP. Fig. 4 illustrates the relationship between the average Q_s and μ_{ref} for the "News" sequence obtained empirically, which can be approximated with a linear function.

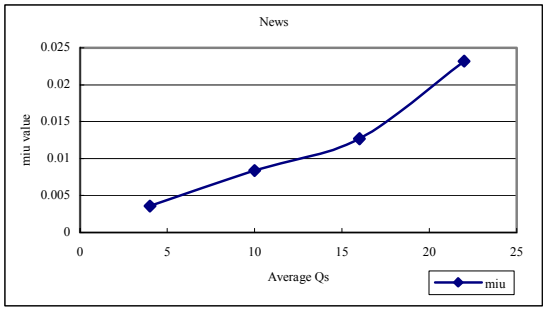


Fig. 4. Effect of Q_s value on μ_{ref} ($Q_{\mathrm{H}} = 22$).

The distortion function $D_i^{\mathrm{rq}}(Q_{s,i})$ can be formulated as a quadratic function as suggested in [8]:

$$D_i^{\mathrm{rq}}(Q_{s,i}) = a_1 Q_{s,i}^2 + a_2 \quad (3)$$

where a_1 and a_2 are model parameters.

According to our experiments, the approximation error between the model in (3) and the actual distortion is lower than 6% on average, meaning that the model in (3) has fairly good accuracy.

In [8], the rate function $R(Q_s)$ for H.26L coding is suggested to be modeled as $R(Q) = b_1 \log_2(b_2 / Q^2)$. By incorporating the effect of quantization step-sizes for the higher and lower bit-rate bitstreams, we may use the following model.

$$R(Q_s) = b_1\left(\log_2 \frac{b_2}{Q_s^2}\right)(Q_{\mathrm{L}} - Q_{\mathrm{H}})^2 = b_1' - b_2' \log_2 Q_s \quad (4)$$

where Q_{H} and Q_{L} are the quantization step-sizes of higher bit-rate and lower bit-rate bitstreams, respectively, Because Q_{H} and Q_{L} are known in advance, their effects can be absorbed into the model parameters b_1' and b_2', by setting $b_1' = b_1(Q_{\mathrm{L}} - Q_{\mathrm{H}})^2 \log_2 b_2$ and $b_2' = 2b_1(Q_{\mathrm{L}} - Q_{\mathrm{H}})^2$.

However, we found that the accuracy of this rate model is not good enough for modeling the bit-rates of secondary SP-frames because the coding of secondary SP-frames is different from that of H.264 P-frames. According to our experiments, we suggest a more accurate model as follows:

$$R(Q_s) = b_1' - b_2' Q_s \log_2 Q_s \quad (5)$$

Table 1 shows the average model error of (5) is only about 3.3-3.5%, whereas that of (4) can be up to 16% for two test sequences. The model in (4), however, leads to difficulty in mathematical manipulations in deriving an optimal close-form solution for (1). Note that, in typical range of Q_s, $\log_2 Q_s \approx Q_s^{1/2}$, we thus can use $Q_s^{3/2}$ to approximate $Q_s \log_2 Q_s$ to make it mathematically tractable while retaining good accuracy for solving (1), leading to the following model.

$$R(Q_s) \cong b_1' - b_2' Q_s^{3/2} \quad (6)$$

Table 1 shows the model error of (6) is only about 4 % compared to the actual bit-rates, which is very close to the model in (5) and much more accurate than that of (4).

Table. 1. Approximation errors of different rate models

	$\log_2 Q_s$	$Q_s \log_2 Q_s$	$Q_s^{3/2}$
News	16.1%	3.5%	4.2%
Stefan	14.3%	3.3%	4.1%

Substituting (2), (3), and (6) into (1), the Lagrange multiplier can be used to convert (1) to the following unconstrained optimization problem as follows:

$$\begin{aligned} f &= \sum_{i=2}^{N_{\mathrm{GOP}}} D_i(Q_{s,i}) + \lambda\left(\sum_{i=2}^{N_{\mathrm{GOP}}} R_i(Q_{s,i}) - R_C\right) \\ &= \sum_{i=2}^{N_{\mathrm{GOP}}} (1+\mu_{ref})^{N_{\mathrm{GOP}}-i}\left(a_1 Q_{s,i}^2 + a_2\right) + \lambda\left[\sum_{i=2}^{N_{\mathrm{GOP}}}\left(b_1 - b_2 Q_{S,i}^{3/2}\right) - R_C\right] \end{aligned} \quad (7)$$

where λ is the Lagrange multiplier.

By setting partial derivatives to zero (i.e.$\partial f/\partial Q_{s,i} = 0$ and $\partial f/\partial \lambda = 0$), we can find the set of $Q_{s,i}$'s which minimizes the cost function in (7) as follows:

$$Q_{s,i} = \left(\frac{N_{\mathrm{GOP}} b_1 - R_C}{b_2 \sum_{i=2}^{N_{\mathrm{GOP}}} \frac{1}{(1+\mu_{\mathrm{ref}})^{3(N_{\mathrm{GOP}}-i)}}}\right)^{2/3} \Big/ (1+\mu_{\mathrm{ref}})^{2(N_{\mathrm{GOP}}-i)} \quad (8)$$

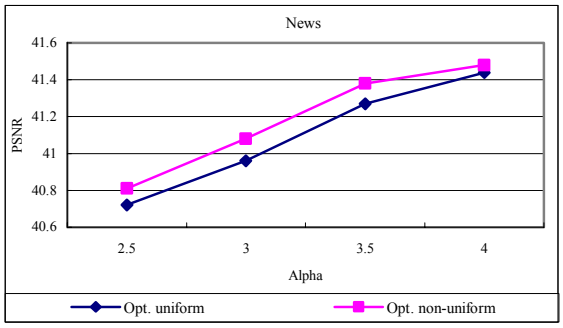
As a result, we can use (8) to find the optimal quantization step-size for each secondary SP-frame in a GOP. A two-pass encoding procedure is adopted in our work. While performing the first-pass encoding, we collect the encoding parameters of all SP-frames within a GOP, including the distortions of primary SP-frames and bit-rates of secondary SP-frames with a few number of Q_s's. We can subsequently use the information to obtain the model parameters of distortion and bit-rate functions using least-squares curve fitting. After estimating the model parameters, we then apply $\mu_{\mathrm{ref}} = 0$ to (8) to obtain the optimal uniform quantization step-size (because $\mu_{\mathrm{ref}} = 0$ means there is no reference relation between every two consecutive frames within a GOP). The optimal uniform quantization step-size is then used to obtain a proper μ_{ref} value according to Fig. 3. Finally we apply the new μ_{ref} to (8) to obtain the optimal quantization step-size for every frame.

IV. Experimental Results

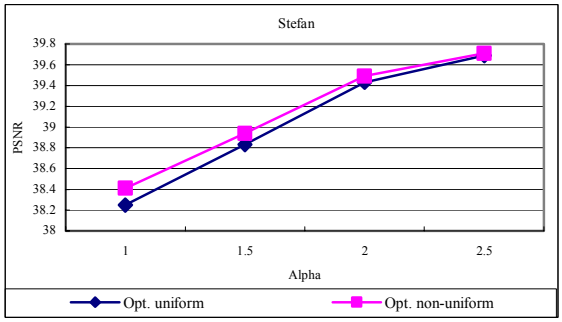
Two QCIF (176×144) test sequences, "News" and "Stefan," are used in our experiments. We use the H.264 reference codec (JM 7.3) to encode the SP-frames with a GOP size of 30 frames and a frame rate of 30 fps. For simplicity of experiments but without loss of generality, we use only two different bit-rate bitstreams encoded with two different fixed quantization step-sizes, $Q_{\mathrm{H}} = 22$ and $Q_{\mathrm{L}} = 28$, for switching. The GOP structur for the two primary bitstreams is IPPP…

Fig. 5 shows the coding performance comparison of proposed R-D optimized uniform and non-uniform quantization schemes under different rate constraints (2.5~4 times of R_{H}). The simulation results show that the average PSNR performance of the optimal non-uniform

quantization scheme is better than that of the optimal uniform one by 0.1-0.15 dB. Fig. 6 shows the frame-by-frame PSNR performance comparison of the two R-D optimized quantization methods. In this figure, the "Upper bound" curve represents the coding performance of higher-quality bitstream and the "Lower bound" curve represents the coding performance of lower-quality bitstream. We can observe from Fig. 6 the different characteristics of the two methods and why the non-uniform method can outperform the uniform one. The uniform quantization method assigns the same Q_s value to all the SP-frames in a GOP, thereby leading to a rather flat quality curve. With the rate and distortion models in (2) and (6), the proposed R-D optimized uniform quantization scheme can achieve the best performance while meeting a given rate constraint accurately using a fixed quantization step-size for all SP-frames in a GOP. The R-D optimized non-uniform quantization scheme, however, applies different quantization step-sizes to different frames according to the distortion models of individual frames and the estimated amounts of error propagation to their subsequent frames due to the quantization distortion. It thus tends to assign finer Q_s values to the former frames and coarser Q_s values to the later frames. Such arrangement will lead to non-uniform picture quality as illustrated in Fig. 6, but can achieve better average performance than the uniform scheme.



(a)

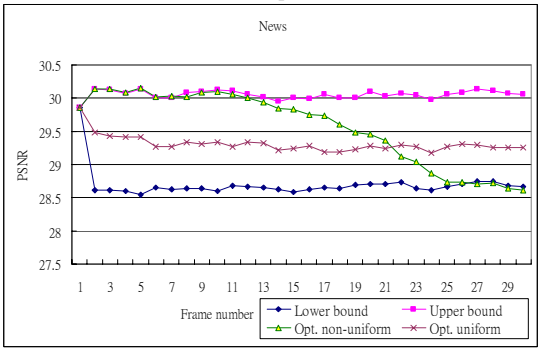


(b)

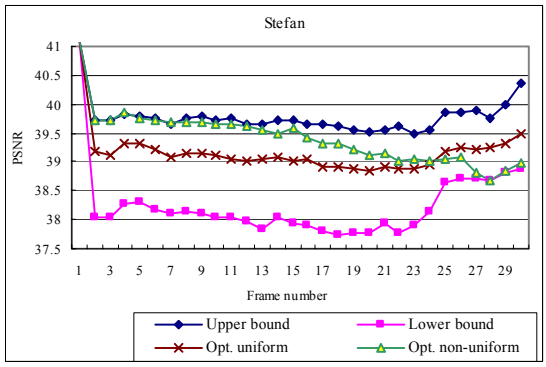
Fig. 5. Average PSNR performance comparison with the proposed R-D optimized uniform and non-uniform rate-control schemes: (a) "News," and (b) "Stefan."

V. CONCLUSION

We proposed efficient rate-distortion optimized quantization schemes to improving the coding efficiency of H.264 primary SP-frames under a given storage constraint on secondary SP-frames. A two-pass encoding procedure is adopted in our work. While performing the first-pass encoding, encoding statistics are collected to estimate the parameters of the distortion and rate models of primary and secondary SP-Frames in a GOP, respectively. With these models, we then use the Lagrange multiplier method to find the optimal set of quantization step-sizes for all frames in a GOP. We have proposed two R-D optimized rate-control schemes: uniform quantization and non-uniform quantization, respectively. Experimental results show that the proposed rate and distortion models are rather accurate and both methods can achieve accurate rate control to meet the storage constraints. The optimal non-uniform quantization scheme can achieve performance improvement of about 0.1-0.15 dB over the uniform quantization schemes.



(a)



(b)

Fig. 6. Frame-by-frame PSNR performance comparison with the proposed R-D optimized uniform and non-uniform rate-control schemes: (a) "News," and (b) "Stefan."

REFERENCES

[1] S.-F. Chang, "Video adaptation: Concepts, technologies, and open issues," *Proc. IEEE*, vol. 93, no. 1, Jan. 2005. (in press)

[2] N. Farber and B. Girod, "Robust H.263 compatible video transmission for mobile access to video servers," in *Proc. Int. Conf. Image Processing*, Oct. 1997, Santa Barbara, CA.

[3] B. Xie and W. Zeng, "Source characteristics based fast bitstream switching," in *Proc. IEEE Int. Conf. Multimedia & Expo*, July 2003.

[4] B. Xie and W. Zeng, "Rate-distortion optimized daynamic bitstream switching for scalable video streaming," in *Proc. IEEE Int. Conf. Multimedia & Expo,* June 2004, Taipei, Taiwan.

[5] M. Karczewicz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol.13, no. 7, pp. 637-644, July 2003.

[6] X. Sun, F. Wu, S. Li, W. Gao, and Y.-Q. Zhang, "Seamless switching of scalable video bitsreams for efficient streaming," *IEEE Trans. Multimedia.*, vol.6, no. 2, pp. 291-303, Apr. 2004.

[7] Y. Yu, Z. Zhou, Y. Wang, and C. W. Chen, "A novel two-pass VBR coding algorithm for fixed-size storage constraint," *IEEE Trans. Circuits Syst. Video Technol.*, vol.11, no. 3, pp. 345-356, Mar. 2001.

[8] T. Wiegand and B. Giord, *Multi-frame Motion-compensated Prediction for Video Transmission*, Kluwer Academic, Boston, USA, 2001.